A CASE STUDY ON SELF-PERCEPTION OF NEW UNDERGRADUATE STUDENTS FOR SOCIAL PROACTIVITY

**[1]GERALDO MAGELA FREITAS DOS SANTOS,
[2]EDILSON FERNEDA, [2]RENATO GUADAGNIN,
[2,3]HÉRCULES ANTONIO DO PRADO**

[1] Centro Universitário do Espírito Santo - UNESC
Av. Fioravante Rossi, 2.930, Colatina, ES, Brazil,
CEP: 29.703-900
E-mail: gemagela@terra.com.br

[2] Universidade Católica de Brasília, Mestrado em Gestão do Conhecimento e da Tecnologia da Informação
SGAN 916, Brasília, DF, Brazil, CEP: 70790-160
E-mails: {eferneda, renatov, hercules} @ucb.br

[3] Embrapa - Brazilian Agricultural Research Corporation,
Parque Estação Biológica - PqEB s/n°, Brasília, DF,
Brazil, CEP 70770-901
E-mail: hercules@embrapa.br

1. INTRODUCTION

The importance of applied studies of pattern recognition techniques relies on the fact that all effort for technology development should target practical results for society. In this sense, largely scrutinized techniques are being applied to solve practical problems in many domains. This paper approaches the problem related to the management of social groups in an education context. In order to adequately deal with the convergent characteristics of a group of students, it is necessary to enlighten some key characteristics. The juvenile leadership expresses a concept of adolescents as rights subject, with power to democratically participate in social change [1]. Juvenile leadership is a way to face violence and alienation among young people [2, 3]. Thus, it is necessary to know the social willingness of the young in a process of education for values.

A case study is reported, in which a self-perception survey of new undergraduate students for social proactivity is associated with socioeconomic features. Data Mining over the students' data base was driven under the guidelines of CRISP-DM. The detected predominance of non-proactivity encourages new studies to assess possible links between this finding and existing socio-cultural features.

2. PROBLEM STATEMENT

According to Maturana and Varela [4], an autopoietic machine has a network of processes of production of components which: *(i)* continuously regenerate the network of processes that produced them; and *(ii)* constitute the machine as a concrete unity in space in which the components exist. The world is not finished and pre-given. Instead of seeing ourselves as fully conditioned beings, we find that there is a space for action and building a better society. Experience is tied to our structure, so that one can not separate the individual history of biological and social actions in relation to the outside world. In this sense, humans are considered as autopoietic experiencing structural, natural, social, cultural, and linguistic relations. By doing this, the baseline to study proactivity is set.

The language makes possible the communication phenomenon with an exchange of meanings in a social and cultural network of interactions. This dynamics takes place inside the social context, in a mutual coupling of a network of reciprocal interactions, which are formed in the so-called third-order units (social and cultural engagement). There are continuity and change processes in socio-cultural orientations that are related to undergraduate study, in accordance with critical analysis of habitus Bourdieusian concept on the social nature of human behavior [5]. Thus one can go beyond a purely objectivist perspective (living conditions, independent of human action) or subjective perspective (human action without considering socio-cultural conditions where action occurs).

Any subject is able both to organize themselves in response to disturbances of physical and socio-cultural environment (reactivity), and also organize this environment (proactive). So, he is structured and structuring.

The language has a fundamental role in social interaction, producing a structural coupling, which carries images, showing different forms of observation and perception of reality by means of symbols (culture, interests, power, beliefs, values, illusions, blind spots). The problem that arises is the identification of proactivity as the possibility of transforming society from the perspective of social inequality.

This paper aims at identifying and analyzing the self-perception of the young

in relation to willingness for social proactivity. It describes the perception of a group of new comers in undergraduate studies on the proactivity in social transformation, from the texts of essays and data from the socioeconomic questionnaire, using a Data Mining technique.

3. MATERIAL AND METHODS

3.1 THE DATA SET

The training dataset was obtained from a socioeconomic questionnaire along with data extracted from essays from an entrance examination. It includes data like schooling and work of the parents, the English knowledge level, the writing grade, and the candidate occupation. Moreover, features that identify the proactiveness were extracted from the candidate writings. For example, the use of the pronouns *I* or *We* when approaching the social inequality, depending on the text construction, may be interpreted as the willing of social proactivity.

3.2 THE MINING PROCESS

Data Mining was driven under the CRISP-DM (Cross Industry Standard Process for Data Mining) guidelines, the well-known method that proposes a set of tasks organized in phases, generic tasks, specialized tasks, and process instances [6]. The core of the method is the model building that was carried out by applying the Combinatorial Neural Model (CNM), a hybrid neural network that represents an alternative to overcome the black box limitation of the Multilayer Perceptron. This is achieved due to a particular characteristic of the model that is built up on a neural network structure along with a symbolic processing [7, 8].

CNM can identify relations among input vectors and output values, performing a symbolic mapping. It uses supervised learning and a feedforward topology with three layers (Fig. 1):
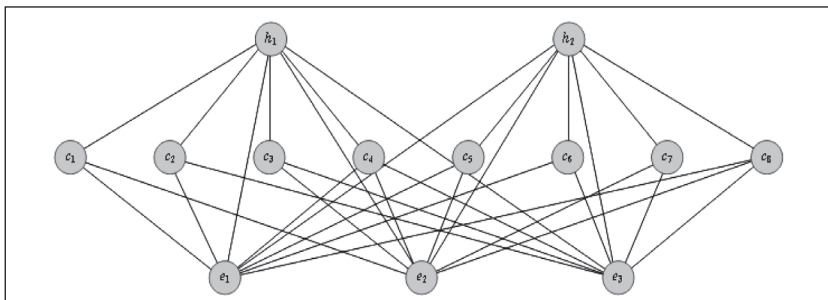

Fig. 1. Example of a combinatorial neural network

(i) the input layer, in which each node corresponds to a triple object-attribute-value that describes a dimension (here called *evidence* and denoted by *e*) in the domain;

(ii) an intermediary or combinatorial layer with neurons connected to one or more neurons from the input layer, representing the logical conjunction AND; and

(iii) the output layer, with one neuron for each possible class (here called *hypothesis* and denoted by *h*), that is connected to one or more neurons in the combinatorial layer by the logical disjunction OR.

The synapses may be inhibitory or excitatory and have assigned a weight between zero (unconnected) and one (fully connected). The network is created as follows: *(i)* one neuron in the input layer for each evidence in the training set; *(ii)* a neuron in the output layer for each class in the training set; and *(iii)* one neuron in the combinatorial layer for each possible combination of evidences (with order ≥ 2) from the input layer. Combinations of order = 1 are connected directly to the output layer. CNM is trained according to the algorithm in Fig. 2.

For the sake of simplicity, without quality loss, we used a constant value 1 for the evidential flow. For CNM training, examples are presented to the input layer, triggering a signal from each neuron matching to the combinatorial layer, having their weights increased. Otherwise, their weights are weakened.

After training, the accumulators associated to each arc in the output layer will belong to the interval $[-c, c]$, where *c* is the number of cases in the training set. After the training process, the network is pruned, based on the accumulators values, as follows: *(i)* remove all arcs arriving to the output layer with accumulators below a threshold specified by the user; and *(ii)* remove all neurons and arcs from the input and combinatorial layers disconnected after the first step.

The relations that remained after pruning are rules that are considered in the application domain.

Two basic metrics are applied during the model generation that allow the evaluation of the resulting rules: Confidence (C) and Support (S). CI

PUNISHMENT_AND_REWARD_LEARNING_RULE

1.Set the initial value of the accumulator in each arc to zero;
2.For each example from the training set, **do**:
    Propagate the evidences from input nodes to the output layer, computing all possible
    combinations of the input nodes involved;
    For each arc arriving to a neuron in the output layer, **do**:
      If the output neuron corresponds to a correct class
      Then backpropagate from this neuron to the input nodes increasing the accumulator of
         each traversed arc by its evidential flow (reward)
      Else backpropagate decreasing the accumulators by its evidential flow (punishment).

Fig. 2. Learning algorithm for CNM.

(Confidence Index) express the degree of cohesion between the premises (the antecedent A) and the conclusion (the consequent C) and indicates the percentage of cases that occurred (antecedent associated with the consequent), compared to the antecedent [9]. Thus

CI =100*(A ∩ C) / A

S indicates the percentage of occurrence of the rule, compared to the consequent. So

S = 100*(A ∩ C) / C

The main problem with CNM is its low performance due to the exponential growing of the combinatorial layer. However, it received many improvements that have turned it feasible as a useful approach for building classifiers [9, 10, 11, 12, 13, 14, 15].

3.3 CNM UTILIZATION DETAILS

The sample in the present study comprises 98 randomly selected students from entrance examination courses, representing 20% of the population. The input data set refers to the answers to the socio-economic questionnaire (objective responses) and to the essays texts that were later transformed into structured data.

Socio-economic questionnaire contains 30 attributes on personal data, past education data, future education data, and family conditions data.

The instrument of analysis of the essays was built by defining the type of knowledge to be discovered in texts, according to Fig. 3.

The phenomena of social inequality included the following possibilities: nationality, richness and poorness, housing, ethnicity, sex, gender, generations, rural and urban, center and periphery, qualification to work, schooling, culture, religion, social class, related to children, teen pregnancy, drugs, violence, murder, and other. Factors responsible for social inequality are: *I*, *we*, *young people*, *society in a general sense*, *civil society*, *politicians*, *government*, *religious authorities*, *elite*, *political parties*, *businessmen*, *non-government organizations*, *financial systems*, *education*, *employers*, *employees*, *birth rates*, *poor people*, *bandits*, *human nature*, and others. Processes or systems that cause

CATEGORY	SUBCATEGORY
Inequality context	Phenomena
	Responsible
	Processes
Changes	Agents
	Processes
Equality horizon	Certainty
	Hope
	Desperation

Fig. 3. Social Inequality and equality

or stimulate social inequality are: neoliberalism, capitalism, *globalization*, *socialism*, *oppression*, *corruption*, *asymmetry of power*, *prejudice*, *moderated politics*, *education*, *media*, *ideology factors*, *public policy*, and others.

Social change agents can be the same as factors that cause social inequality. Processes and dynamics of social change include *fight*, *radicalization*, *history*, *proactivity*, *public policy*, *education*, *law*, *science*, *competition*, *ethical-political instances*, *values*, *religion*, *ecology*, *election*, and other.

The target variable or attribute classifier was proactivity, understood operationally as expressions contained in the texts of essays concerning social change, which had as action subject or action intention the *I* or *we*. Thus, this attribute has four possibilities as domain description: *(i) pro-active young*, *(ii) non pro-active young*, *(iii) pro-active adult* or *(iv) non pro-active adult*. Attribute proactivity were derived from attribute *I* or *we*, along with attribute age. Entrance examination participants were considered *young* if s/he was under 25 and *adult* otherwise. The data obtained were included into the database of socio-economic questionnaire. Attributes with frequency under 10 and over 96 were excluded, as well low discriminative skill.

4. RESULTS

The essays were anonymously analyzed by three teachers with master degree in Arts and experience in correcting essays, based on the previously described instrument (Fig. 3). The Pearson correlation between the classifications obtained by the three professors who evaluated the essays was 0.75 between teachers A and B, 0.80 between teachers A and C, and 0.85 between the teachers B and C. Therefore, we obtained a positive correlation, above moderate and below strong, according to Levin [16]. The final value assumed for the target variable was defined on the basis of a majority criteria, i.e., when at least two readers agreed with the value.

On average, the number of attributes found by essay was 7 among 96 attributes, and 5.3 among the attributes potentially significant, i.e., with a minimum 10% support. The categories with support less than 10% were excluded from the training base. Regarding the 96 attributes of the instrument of initial analysis of essays, 19 of them were potentially significant, representing a recovery of about 20%. The database for mining has 52 attributes, 33 of the socio-economic questionnaire and 19 of the instrument of essay analysis. The scores achieved by entrance examination participants in sample ranged from 0-17 on a 0-20 points scale. The average was 9.12 and the standard deviation was 3.34.

It was selected 71 rules from a total of 1640 generated. The selected rules were those that reached a CI value not less than 90% and a level of support higher than 10%. Fig. 4 shows some of such rules.

Premise	Conclusion	CI	Cases	S
The applicant has provided only a entrance examination AND The father has regular work AND Your writing grade ∈ [mean, mean + standard deviation]	Non Proactive Young	100.00 %	7	14.29 %
The father's schooling is to complete high school AND The mother works regularly AND The protagonist of social change is the Government	Non Proactive Young	100.00%	7	14.29 %
The candidate is not a paid job AND Their level of understanding of English is regular AND Your note writing ∈ [mean, mean + standard deviation]	Non Proactive Young	100.00%	7	14.29 %

Fig. 4. Example of selected models from the MN

5. DISCUSSION

The predominant profile in the sample is the non pro-active young, to whom social change depends upon the actions of other social agents, rather than from them. All the standards listed below that describe this non proactive young, come from the selected rules (those with CI ≥ 90% and S > 10%). The non pro-active young people assign the responsibility for proactiveness in social change to the government and to the non profession oriented courses he attended in private schools.

He concluded high school less than three years ago, he has already participated in an entrance examination, he has a regular understanding of English, he belong to families with up to five dependents in family income, his father has undergraduate studies and works regularly, and his mother has high school studies and works regularly.

For non proactive adults, the CNM describes combinations of the following features:
- 100% CI and 26.92% support:
– humanities as entrance examination area;
– enrolled in the Social Services, and
– finished high school for over three years.
- 80% CI and 44.44% support:
– married,
– the mother is retired, and
– the father is educated only up to 4th grade of elementary school.

- •   80% CI and 15.38% support:
– attended a profession oriented high school;
– high school studies in public school;
– finished high school for over three years;
– attended an entrance examination course for one year;
– married;
– monthly family income is 1-3 times the minimum wage;
– his father is retired or died without leaving a pension;
– his mother has finished elementary school, while his father completed the primary school;
– follows another religion (other than those mentioned in the questionnaire);
– already participated in two entrance examinations;
– points out difference between social classes as a phenomenon of social inequality;
– to reduce social inequality, suggests paths to education and public policy.

Young students that were classified as proactive have an employed mother in a small business. Here, CI is less than 90% (75% CI and 21.43% support). As the candidates were classified as not the most pro-active, the rules generated by CNM revealed patterns just for that category.

The most highlighted social agent was *government*. The most prominent mechanisms toward social changes were public policies, especially education and employment.

6. CONCLUSION

It was possible to detect two distinct profiles of entrance examination participants: the *non proactive young* and the *non proactive adult*, since non proactivity was the main feature in the studied group. The generated models effectively subsidize the decision making in the other dimensions of school management, such as: administrative, financial, publicity, advertising, communication, and marketing. Achieved results reflect the perceptions of young people, according to their social, historical, and cultural conditions, as a starting point for a joint educational effort.

Some issues need to be further investigated:

(i) The relationship between age, religious belief, education for human values and proactivity in social change;

(ii) The adoption of a taxonomy for the social proactivity, beyond the simple dualism proactive/non proactive;

(iii) The inclusion of other characteristics (being a lazy or an affirmative

person, for example) based on the handwriting style can be fruitful to enrich the understanding of individual classes.

Results show that current approach is useful for educational institutions as well as for any other domain where one can subjectively gather data concerning personal opinions.

REFERENCES

1. STAMATO M.I. C. Juvenile Protagonism: A Social-Historical Praxis to Prepare to Citizenship. Proceedings of the XV Encontro Nacional da Associação Brasileira de Psicologia Social (ENABRAP-SO) – Mesa Redonda ED MR070 - Formação Humana e Profissional, Maceió, Brazil, 2009. Available in: http://abrapso.org.br/siteprincipal/images/Anais_XVENABRAPSO/389.%20protagonismo%20juvenil.pdf. (in Portuguese).

2. MITRULISE. Experiments for Innovation in High Scholl Level. Cadernos de Pesquisa, São Paulo, nº 116, p. 217-244, jul. 2002. Available in: http://www.scielo.br/pdf/cp/n116/14404.pdf. (in Portuguese).

3. ANDRADE S.F.S., NUNES C.A.A. Physics Learning and Juvenile Protagonism. Proceedings of the XVI Simpósio Nacional de Ensino de Física, CEFET-RJ, Rio de Janeiro, Brazil, 2005. Available in: http://www.sbf1.sbfisica.org.br/eventos/snef/xvi/cd/resumos/T0592-1.pdf. (in Portuguese).

4. MATURANA H., VARELA F. ([1st edition 1973] 1980). Autopoiesis and Cognition: the Realization of the Living. Robert S. Cohen and Marx W. Wartofsky (Eds.), Boston Studies in the Philosophy of Science 42. Dordrecht: D. Reidel Publishing Co.

5. CASANOVA J.L.S. Social Orientations – a Critical and Operative Approach to the Concept of *Habitus*. In: Proceedings of the V Congresso da Associação Portuguesa de Sociologia. Portugal, 2007. Available in: http://www.aps.pt/cms/docs_prv/docs/DPR4628fe7bb4abb_1.pdf. (in Portuguese)

6. CHAPMAN P. et al. CRISP-DM 1.0 Step-by-step data mining guide. SPSS. 2000. Available in: http:// www.crisp-dm.org/CRISPWP-0800.pdf.

7. MACHADO R.J., Rocha A.F. Handling knowledge in high order neural networks: the combinatorial neural network. Rio de Janeiro, Brazil: IBM Rio Scientific Center, Brazil, 1989. (Technical Report CCR076).

8. HAYKIN S. *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 1999.

9. FELDENS M.A., CASTILHO J.M.V. Data Mining with the

Combinatorial Rule Model: An Application in a Health-Care Relational Database, *Proceedings of the XXIII Latin-American Conference on Informatics* (CLEI), Valparaiso, Chile, 1997.

10. MACHADO R.J., CARNEIRO W., NEVES P.A. Learning in the combinatorial neural model, *IEEE Transactions on Neural Networks*, Vol. 9, p. 831-847, 1998.

11. BECKENKAMP F.G., PREE W., FELDENS M.A. Optimizations of the Combinatorial Neural Model, Proceedings of the 5th Brazilian Symposium on Neural Networks, p. 49, 1998.

12. PRADO H.A., FRIGERI S.R., ENGEL P.M. A Parsimonious Generation of Combinatorial Neural Model. IV Argentine Congress on Computer Science (CACIC'98), Neuquén, Argentina, 1998.

13. PRADO H.A., MACHADO K.F., FRIGERI S.R., Engel P.M. Accuracy Tuning on Combinatorial Neural Model. In: N. Zhong and L. Zhou, Eds. *Proceedings of the Third Pacific-Asia Conference on Methodologies For Knowledge Discovery and Data Mining*, p. 247-251, LNCS, Vol. 1574, Springer-Verlag, London, 1999.

14. PRADO H.A., MACHADO K.F., ENGEL P.M. Alleviating the complexity of the Combinatorial Neural Model using a committee machine. *Proceedings of the International Conference on Data Mining*, Cambridge, UK: WIT Press, 2000.

15. NOIVO R., PRADO H.A., LADEIRA M. Yet Another Optimization of the Combinatorial Neural Model. *Proceedings of the XXX Latin-American Conference on Informatics* (CLEI), p.706-711, Arequipa, Chile, 2004.

16. LEVIN J. Elementary Statistics in Social Research (11th Edition), Allyn & Bacon 11ª., 2009.

Geraldo Magela Freitas dos Santos. Degree in Pedagogy from the Don Bosco School of São João del Rei. Specialist in School Supervision from the Catholic University of Minas Gerais. Master in Education from the Salesian Polytechnic University, Quito. Ecuador. Retired Professor from the Federal University of São João del Rei, MG. He is currently Director of ASSESC - Advisory and Consultancy in Education and Information Systems and Academic Advisor of the Rectorate of UNESC - University Center of Espírito Santo, Colatina, Brazil.

Edilson Ferneda. He holds a Computer Technology undergrad by Technology Institute of Aeronautics (1979), Master in Computer Science, Federal University of Paraíba (1988), and Ph.D. in Computer Science by the Laboratoire d'Informatique, Robotique et Microélectronique de Montpellier (1992). He is currently professor at the Graduate Program in Knowledge and Information Technology Management Catholic University of Brasilia. He has experience in Computer Science, with emphasis on applied artificial intelligence and knowledge management.

Renato Guadagnin. Mechanical Engineer by Polytechnic School of Federal University of Rio de Janeiro (1969), Master of Systems Analysis by National Institute for Space Research (1972), and Ph.D. by German University of Administrative Sciences, Speyer (1984). He is retired professor from University of Brasilia. Presently he is professor at Catholic University of Brasilia. His scientific works concern: knowledge management, information technology management, artificial intelligence, image analysis and visualization.

Hércules Antonio do Prado. He holds a Data Processing undergrad by Federal University of São Carlos (1976), a M.Sc. in Computer Science by Federal University of Rio de Janeiro (1987), and a D.Sc. in Computer Science by Federal University of Rio Grande do Su (2003)l. He joined the Brazilian Enterprise for Agricultural Research in 1984, developing research on computational methods applied to agricultural research. In 1992 he joined the Catholic University of Brasilia as a lecturer in Computer Science and, later, as a researcher in the Graduate Program in Knowledge and Information Technology Management. His research interests includes data mining, knowledge management, organizational learning, and competitive intelligence.